数理解析研究所講究録
第 1621 巻 2009 年 135-152

Tweedie一般化線形モデルを用いたクイーンズランド州の降水量データの解析

統計数理研究所　大西 俊郎 (Toshio Ohnishi)
The Institute of Statistical Mathematics
サンシャインコースト大学　Peter K Dunn
University of the Sunshine Coast

Abstract

We investigate a logarithmic-link generalized linear model, whose underlying sampling density is in an exponential family distribution with power variance function. The multiple-strata case is studied with stratum-dependent intercepts and a common slope. We prove that there exists a conjugate prior density on the intercept parameter, and the conjugate analysis is discussed. An estimation procedure is given, which includes the optimal estimating function of the parameters other than the intercept, and an empirical Baysian estimation of the hyper-parameters of the prior density. As an example, rainfall data for Queensland, Australia, is analyzed.

Key Words: common slope, conjugate analysis, empirical Bayesian estimation, estimating function, generalized linear model, logarithmic link, Pythagorean relationship, posterior mode, Tweedie distribution

1 Introduction

The generalized linear model (GLM, McCullagh & Nelder [14]) plays an important role in data analysis, enjoying wide application in fields such as insurance, climatology, economics and biostatistics. Their popularity is partially because GLMs are based on the exponential dispersion model (EDM) family of distributions, which includes common distributions such as the Normal, binomial, Poisson and gamma distributions as special cases. In addition, inference for the GLM has a minimax property: an exponential family distribution minimizes the Fisher information for the mean parameter under a given mean–variance relationship (Tsubaki [21]; Ohnishi & Tsubaki [15]).

If the response variable Y follows an EDM distribution with mean μ, the variance is $\mathrm{var}[Y] = V(\mu)/\tau$, where $\tau > 0$ is a precision parameter, and $V(\mu)$ is the variance function, some function of the mean.

A special subset of EDMs are the Tweedie distributions, with variance function $V(\mu) = \mu^p$ for some p. Special cases include the Normal ($p = 0$), Poisson ($p = 1$ with $\tau = 1$), gamma ($p = 2$) and inverse Gaussian ($p = 3$) distributions. Tweedie distributions exist for all $p \notin (0, 1)$.

The Tweedie distributions with $p \in (1, 2)$ are of interest here. These distributions have support on the non-negative reals, and are continuous for $Y > 0$ with a positive probability $\mathrm{P}(Y = 0)$.

Until recently, Tweedie distributions (apart from the four special cases identified above) were rarely used since their distributions, and hence likelihood functions, cannot be written in closed form. Recent advances have produced accurate numerical methods for evaluating the density functions (Dunn & Smyth [3, 4]).

Although Bayesian analysis for GLMs is well established (for example, see Gelman et al. [8]), little is known about Bayesian analysis using Tweedie distributions. A conjugate analysis of a particular Bayesian GLM is addressed in this paper. Chen & Ibrahim [1] discuss the conjugate analysis of GLMs for both the intercept and slope, using the power prior [10], but their analysis depends on the covariates.

We assume a conjugate prior density on the intercept parameters, and develop the conjugate analysis under this prior density. Unlike Chen & Ibrahim [1], our priors are assumed on the intercept only, and our priors do not depend on the covariates, but use hyper-parameters. Also, the situation under study assumes the data consist of K strata, with a common slope of interest, but separate intercepts. The Bayesian approach is known to perform well when the dimension of the parameter space is high; the James–Stein estimator [11] is a well-known example (see Efron [5]). In our model K is assumed to be large, say more than 35.

After first introducing GLMs and EDMs (Sect. 2), and Tweedie EDMs in particular (Sect. 3), the likelihood function for the scenario under study is developed (Sect. 4), followed by a conjugate analysis of the intercept parameter (Sect. 5). The optimal estimating function is established (Sect. 6) as well as an estimation procedure (Sect. 7). The results are then demonstrated using an example (Sect. 8).

2 Generalized linear models, the exponential family and location–dispersion models

The EDM family of distributions have probability functions

$$f(y;\ \mu,\ \tau) = \exp\left[\tau\{c(\mu)y - M(\mu)\}\right] a(y;\ \tau), \tag{1}$$

where μ is the mean, and $\tau > 0$ is the precision parameter. The known function $c(\mu)$ is the canonical parameter; the known function $M(\mu)$, when written in terms of the canonical parameter, is the cumulant function; $a(y;\ \tau)$ is the supporting measure. The variance of Y is $\mathrm{var}[Y] = V(\mu)/\tau$ where the variance function $V(\mu)$ is

$$V(\mu) = \left\{\frac{dc(\mu)}{d\mu}\right\}^{-1},$$

which uniquely identifies the distribution in the EDM family (Jørgensen [12, §2.3.1]). Use the notation $\mathrm{ED}(\mu, \tau)$ to denote a random variable Y has the EDM distribution in (1). As indicated earlier, common distributions such as the Normal, binomial, Poisson and gamma distributions are in the EDM family. EDMs are important as they are the response distributions for GLMs.

Closely related to the EDM family is the location–dispersion family (Jørgensen [12]). Distributions in this family have the form

$$p(y - \mu, \tau) = \exp\left\{-\tau d(y - \mu) + b(\tau)\right\}, \tag{2}$$

for a normalizing constant $\exp\{b(\tau)\}$, where μ and $\tau > 0$ are the location (but not the mean) and precision parameters respectively. The function $d(t)$ is called the unit deviance function, where $d(0) = 0$, and $d(t) > 0$ when $t \neq 0$; that is, $d(t)$ is a distance measure.

GLMs consist of two components (McCullagh and Nelder [14]):

1. The response variable, Y_i, follows an EDM family distribution, with mean μ and precision parameter τ such that $Y_i \sim \mathrm{ED}(\mu_i, \tau w_i)$ where $w_i > 0$ are known prior weights; and
2. The expected values of the Y_i, say μ_i, are related to the covariates $\mathbf{x}_i$ through a monotonic, differentiable link function $h(\cdot)$ so that $h(\mu_i) = \alpha + \mathbf{x}_i^T\beta$, where β is a vector of unknown regression coefficients, and α represents a constant term.

Often the linear component, or linear predictor, $\alpha + \mathbf{x}_i^T\beta$ is denoted by η_i, when $h(\mu_i) = \eta_i = \alpha + \mathbf{x}_i^T\beta$.

3 Tweedie distributions

The EDMs with variance function $V(\mu) = \mu^p$ for some real p are of interest here. For these EDMs,

$$f(y;\, \mu,\, \tau,\, p) = \exp\left[\tau\{c(\mu,\, p)y - M(\mu,\, p)\}\right] a(y;\, \tau,\, p), \tag{3}$$

with

$$V(\mu) = \left\{\frac{\partial c(\mu,\, p)}{\partial \mu}\right\}^{-1} = \mu^p.$$

These EDM distributions are called the Tweedie distributions by Jørgensen [12] in honor of Tweedie [22] . Use the notation $\mathrm{ED}_p(\mu, \tau)$ to denote a random variable Y follows the Tweedie EDM in (3).

Jørgensen [12] shows distributions exist for all values of $p \notin (0,1)$. The Tweedie family includes important distributions, such as the Normal ($p = 0$), the Poisson ($p = 1$ with $\tau = 1$), the gamma ($p = 2$) and the inverse Gaussian ($p = 3$) distributions. The binomial distribution is a notable exception.

When $1 < p < 2$, the density (3) can be represented as a Poisson sum of gamma distributions, and is sometimes called the Poisson-gamma distribution. Suppose N random variables X_i (for $i = 1, \ldots, N$) are observed, where N follows the Poission distribution $\mathrm{Po}(m)$ with mean (and variance) m. Also, suppose each random variable X_i follows a gamma distribution with shape parameter θ and scale parameter λ such that the mean is $\theta\lambda$ and variance $\theta\lambda^2$. Then the distribution of $\sum_{i=1}^N X_i$, where $N, X_1, X_2, \ldots$ are mutually independent, corresponds to an $\mathrm{ED}_p(\mu, \tau)$ distribution where

$$m = \tau\frac{\mu^{2-p}}{2-p}, \quad \lambda = \frac{(p-1)\mu^{p-1}}{\tau}, \quad \theta = \frac{2-p}{p-1}.$$

The probabilty function for the Tweedie distribution with a power parameter $p \in (1,\, 2)$ is

$$f(y; \mu, \tau, p) = \exp\left\{-\tau\left(-\frac{\mu^{1-p}}{1-p}y + \frac{\mu^{2-p}}{2-p}\right)\right\} a(y; \tau, p) \tag{4}$$

for $y > 0$ (Jørgensen [12, Chapter 4]), where

$$a(y; \tau, p) = \frac{1}{y}\sum_{j=1}^{\infty} \frac{\tau^j \left\{\frac{1}{2-p}\left(\frac{\tau y}{p-1}\right)^{(2-p)/(p-1)}\right\}^j}{\Gamma((2-p)j/(p-1))j!}. \tag{5}$$

The Tweedie distribution with $1 \leq p < 2$ has the positive probability of zero

$$\mathrm{P}(Y = 0) = \exp\left(-\tau\frac{\mu^{2-p}}{2-p}\right).$$

The normalizing constant $a(y; \tau, p)$ cannot be written in closed form apart from the four special cases indicated earlier. To evaluate $a(y; \tau, p)$, many authors directly evaluate the summation (5), but Dunn & Smyth [3] presented the first rigorous study of the series expansion, and found it is not possible to accurately evaluate this expansion for all parts of the parameter space. Dunn & Smyth [4] then developed a method for inverting the simple form of the moment generating function of the Tweedie densities, producing accurate computations in the parts of the parameter space where the series is not accuracate. These algorithms are implemented in the **tweedie** package (Dunn [2]) for the R statistical environment [18], and we use these programs to derive our numerical results.

In Tweedie GLMs with $p > 1$, a logarithm link function is commonly used, since it ensures $\mu > 0$ as required.

For later convenience, write the Tweedie density in (4) as

$$f(y;\, \mu,\, \tau,\, p) = \exp\left[-\tau\{u(\mu;\, 2-p) - yu(\mu; 1-p)\}\right]\, \tilde{a}(y;\, \tau,\, p), \tag{6}$$

for $\tilde{a}(y;\, \tau,\, p) = \exp\left[-\tau\{-(1-p)^{-1}y + (2-p)^{-1}\}\right] a(y;\, \tau,\, p)$, where

$$u(t;\, \kappa) = \begin{cases} \log t & \text{for } \kappa = 0, \\ \dfrac{t^{\kappa}-1}{\kappa} & \text{otherwise.} \end{cases} \tag{7}$$

In this form, u is continuous in κ, and is equivalent to the log-limit form used by Dunn & Smyth [3]. The function $u(t;\kappa)$ proves crucial later, based on the following properties. The proof is a straightforward calculation, and is omitted.

Lemma 1 *Consider a function $u(t;\kappa)$ as defined in (7).*

(i) For any s, t and κ, $u(st;\, \kappa) = t^{\kappa}u(s;\, \kappa) + u(t;\, \kappa)$.

(ii) Suppose κ and ν are non-negative and $\kappa + \nu > 0$. Then, for any $q > 0$ and $r > 0$

$$\begin{aligned} &qu(t;\, \kappa) - ru(t;\, -\nu) \\ &= \delta_*\{u(t/t_*;\, \kappa) - u(t/t_*;\, -\nu)\} + qu(t_*;\, \kappa) - ru(t_*;\, -\nu), \end{aligned}$$

where

$$t_* = (r/q)^{\frac{1}{\kappa+\nu}} \quad \text{and} \quad \delta_* = q^{\frac{\nu}{\kappa+\nu}} r^{\frac{\kappa}{\kappa+\nu}}.$$

4 The likelihood function for the Tweedie glm

Although an extension to the vector slope parameter is straightforward, we now focus on the scalar case for simplicity.

In this paper, interest is in the logarithmic link Tweedie GLM with linear predictor. Suppose Y_i is distributed according to $\mathrm{ED}_p(\mu_i, \tau)$ with $\log \mu_i = \alpha + \beta x_i$ $(1 \le i \le n)$; then using the expression (6), the density function is given as

$$\begin{aligned} \prod_{i=1}^{n} f(y_i; e^{\alpha+\beta x_i}, \tau, p) = \exp&\left[-\tau \sum_{i=1}^{n}\{u(e^{\alpha+\beta x_i}; 2-p) - y_i u(e^{\alpha+\beta x_i}; 1-p)\}\right] \\ &\times \prod_{i=1}^{n} \tilde{a}(y_i;\, \tau,\, p). \end{aligned} \tag{8}$$

For use later, use this density function to form the likelihood function, and re-write separating α and β, as shown in the following Proposition.

Proposition 1 *The Tweedie likelihood in (8) can be written as*

$$\begin{aligned} \prod_{i=1}^{n} f(y_i; e^{\alpha+\beta x_i}, \tau, p) = \exp&\left[-n\tau\{A(\beta,p)u(e^{\alpha}; 2-p) - B(\beta,p)u(e^{\alpha}; 1-p)\}\right] \\ &\times \prod_{i=1}^{n} f(y_i; e^{\beta x_i}, \tau, p). \end{aligned} \tag{9}$$

where

$$A(\beta,p) = \frac{1}{n}\sum_{i=1}^{n} e^{(2-p)\beta x_i} \quad \textit{and} \quad B(\beta,p) = \frac{1}{n}\sum_{i=1}^{n} y_i e^{(2-p)\beta x_i}.$$

As well as noting the separation between α and β, note the symmetry between the two components involving $(1-p)$ and $(2-p)$ in (9). This proves useful when the conjugate analysis of the intercept α is studied in Sect. 5.

Proof Apply Lemma 1(i) to the likelihood function based on the density function in (8); the summation term in the exponent of the likelihood function becomes

$$\begin{aligned}
&\sum_{i=1}^{n} \{u(e^{\alpha+\beta x_i}; 2-p) - y_i u(e^{\alpha+\beta x_i}; 1-p)\} \\
&= \sum_{i=1}^{n} \Big\{ e^{(2-p)\beta x_i} u(e^{\alpha}; 2-p) - y_i e^{(1-p)\beta x_i} u(e^{\alpha}; 1-p) \\
&\quad + u(e^{\beta x_i}; 2-p) - y_i u(e^{\beta x_i}, 1-p)\Big\} \\
&= n\{A(\beta,p)u(e^{\alpha}; 2-p) - B(\beta,p)u(e^{\alpha}; 1-p)\} \\
&\quad + \sum_{i=1}^{n} \{u(e^{\beta x_i}; 2-p) - y_i u(e^{\beta x_i}; 1-p)\}.
\end{aligned}$$

From (8), the middle expression is equivalent to the summation term in the exponent of $\prod f(y_i; e^{\beta x_i}, \tau, p)$; then using the given definitions of $A(\beta,p)$ and $B(\beta,p)$, the result follows. □

Now, we consider a specific Tweedie GLM using a logarithmic link function, where the data consist of K strata with n_k observations in stratum k. For each stratum, assume separate intercepts α_k, but common slope β for covariate x_{ki}; primary interest is in the slope β.

The density function is

$$\begin{aligned}
f(\boldsymbol{y};\, \boldsymbol{\alpha},\, \beta,\, \tau,\, p) &= \prod_{k=1}^{K} f(\boldsymbol{y}_k;\, \alpha_k,\, \beta,\, \tau,\, p) \\
&= \prod_{k=1}^{K}\prod_{i=1}^{n_k} f(y_{ki};\, \mu_{ki},\, \tau,\, p),
\end{aligned} \tag{10}$$

where $\log \mu_{ki} = \alpha_k + \beta x_{ki}$. In the above, $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_K)^T$ with $\boldsymbol{y}_k = (y_{k1}, \ldots, y_{kn_k})^T$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)^T$ is the intercept parameter vector, β is the common slope parameter, and x_{ki}s are the covariates.

To simplify the notation, we define several quantities. We extend $A(\beta,\, p)$ and $B(\beta,\, p)$ in Proposition 1 to the multi-strata case as

$$A_k(\beta,\, p) = \frac{1}{n_k}\sum_{i=1}^{n_k} e^{(2-p)\beta x_{ki}} \quad \text{and} \quad B_k(\beta,\, p) = \frac{1}{n_k}\sum_{i=1}^{n_k} y_{ki} e^{(1-p)\beta x_{ki}}, \tag{11}$$

respectively. We also introduce the following quantities:

$$\begin{aligned}
\hat{\alpha}_{\mathrm{M}k} &= \hat{\alpha}_{\mathrm{M}k}(\beta,\, p) = \log\frac{B_k(\beta,\, p)}{A_k(\beta,\, p)}, \\
\delta_{\mathrm{M}k} &= \delta_{\mathrm{M}k}(\beta,\, \tau,\, p) = \tau\{A_k(\beta,\, p)\}^{p-1}\{B_k(\beta,\, p)\}^{2-p}.
\end{aligned}$$

The former will be shown to be the maximum likelihood estimator (MLE) for α given (β, p).

Interestingly, the likelihood function with respect to the intercept parameter α is similar to that of the location–dispersion family, as shown in the following proposition.

Proposition 2 *The likelihood function corresponding to* (10) *is*

$$f(\boldsymbol{y};\, \boldsymbol{\alpha},\, \beta,\, \tau,\, p) = f(\boldsymbol{y};\, \mathbf{0},\, \beta,\, \tau,\, p) \exp\left[-\sum_{k=1}^{K} \delta_{\mathrm{M}k} n_k L(\alpha_k - \hat{\alpha}_{\mathrm{M}k};\, p)\right] \times \prod_{k=1}^{K} \exp\{-n_k C_k(\beta,\, \tau,\, p)\}, \tag{12}$$

where $\mathbf{0}$ *is the* K*-dimensional zero vector,*

$$L(t;\, p) = u(e^t;\, 2-p) - u(e^t;\, 1-p), \tag{13}$$

$$C_k(\beta,\, \tau,\, p) = \tau\{A_k(\beta,\, p) u(e^{\hat{\alpha}_{\mathrm{M}k}};\, 2-p) - B_k(\beta,\, p) u(e^{\hat{\alpha}_{\mathrm{M}k}};\, 1-p)\}.$$

Notice the form of (12) is like that of the location–dispersion family (2), where $\tau \equiv \delta_{\mathrm{M}_k} n_k$ and $d(y-\mu) \equiv L(\alpha_k - \hat{\alpha}_{\mathrm{M}k};\, p)$. Note that $L(t;\, p) > 0$ for $t \neq 0$, and $L(t;\, p) = 0$ if $t = 0$, as required for a deviance function.

Proof As an application of Proposition 1 to the kth stratum, we have

$$\begin{aligned} &f(\boldsymbol{y}_k;\, \alpha_k,\, \beta,\, \tau,\, p) \\ &= \exp\left[-\tau n_k \left\{A_k(\beta,\, p)\, u(e^{\alpha_k};\, 2-p) - B_k(\beta,\, p)\, u(e^{\alpha_k};\, 1-p)\right\}\right] \times \\ &\quad f(\boldsymbol{y}_k;\, 0,\, \beta,\, \tau,\, p). \end{aligned} \tag{14}$$

We apply Lemma 1(ii) to the linear combination of the $u(\cdot;\cdot)$ terms in the exponent. Recalling the definitions of $L(t;\, p)$ and $C_k(\beta,\, \tau,\, p)$, we see that

$$\begin{aligned} &A_k(\beta,\, p) u(e^{\alpha_k};\, 2-p) - B_k(\beta,\, p) u(e^{\alpha_k};\, 1-p) \\ &= \frac{1}{\tau}\{\delta_{\mathrm{M}k} L(\alpha - \hat{\alpha}_{\mathrm{M}};\, p) + C_k(\beta,\, \tau,\, p)\}. \end{aligned} \tag{15}$$

Combining (14) and (15), the likelihood function for stratum k is

$$\begin{aligned} &f(\boldsymbol{y}_k;\, \alpha_k,\, \beta,\, \tau,\, p) \\ &= f(\boldsymbol{y}_k;\, 0,\, \beta,\, \tau,\, p) \exp\{-\delta_{\mathrm{M}k} n_k L(\alpha_k - \hat{\alpha}_{\mathrm{M}k};\, p) - n_k C_k(\beta,\, \tau,\, p)\}, \end{aligned} \tag{16}$$

which completes the proof. □

The function $L(t;\, p)$ is used as a loss function in the conjugate analysis discussed in Sect. 5. Proposition 2 shows that $\hat{\alpha}_{\mathrm{M}k}(\beta,\, p)$ is the MLE for α given β and p.

5 Conjugate analysis of the intercept

Motivated by the result of Proposition 2, assume the following prior density on α_k, which is in the location–dispersion family (2):

$$\pi(\alpha_k - \alpha_0;\, p,\, \delta n_k) = \frac{1}{K(p,\, \delta n_k)} \exp\{-\delta n_k L(\alpha_k - \alpha_0;\, p)\}. \tag{17}$$

Here α_0 and $\delta > 0$ are hyper-parameters, $L(t; p)$ is the deviance function defined in (13), and

$$K(p, t) = \int_{-\infty}^{\infty} \exp\{-tL(s; p)\}\, ds \tag{18}$$

is the normalizing constant. When $p = 1$ or $p = 2$, the density (17) is a log-transformed gamma density.

The prior density (17) may be derived in two different ways.

1. The first is based on the likelihood approach, related to the notion of the power prior proposed by Ibrahim & Chen [10]. To see this, consider (16) as a likelihood function of α_k, supposing the other parameters are known. Replace $\hat{\alpha}_{Mk}$ and δ_{Mk} with the hyper-parameters α_0 and δ, respectively, and the assumed prior density is obtained.
2. The second is an application of the method in Yanagimoto & Ohnishi [23]. The Kullback-Leibler divergence between two Tweedie densities is

$$\mathrm{KL}\big(f(y; \mu_1, \tau, p), f(y; \mu_2, \tau, p)\big) = \tau\mu_1^{2-p} L(\log\mu_2 - \log\mu_1; p).$$

Thus the Kullback-Leibler divergence from model $f(\boldsymbol{y}_k; \alpha_{k1}, \beta, \tau, p)$ to $f(\boldsymbol{y}_k; \alpha_{k2}, \beta, \tau, p)$ is

$$\mathrm{KL}(\alpha_{k1}, \alpha_{k2}; \beta, \tau, p) = \tau L(\alpha_{k2} - \alpha_{k1}; p) \sum_{i=1}^{n_k} e^{(2-p)(\alpha_{k1} + \beta x_{ki})}.$$

Consider the prior density proportional to $\exp\left\{-\tilde{\delta}\mathrm{KL}(\alpha_0, \alpha_k; \beta, \tau, p)\right\}$. Substitution of $\tilde{\delta}\tau \sum e^{(2-p)(\alpha_0 + \beta x_{ki})}$ with δn_k gives the assumed prior density.

Prove the conjugacy of the assumed prior density (17) using Lemma 1.

Proposition 3 *The posterior density corresponding to the prior density* (17) *under the sampling density* $f(\boldsymbol{y}_k; \alpha_k, \beta, \tau, p)$ *is* $\pi\big(\alpha_k - \hat{\alpha}_{Bk}; p, \delta_{Bk} n_k\big)$, *where*

$$\begin{aligned}
\hat{\alpha}_{Bk} &= \hat{\alpha}_{Bk}(\beta, \tau, p, \alpha_0, \delta) = \log \frac{\tau B_k(\beta, p) + \delta e^{-(1-p)\alpha_0}}{\tau A_k(\beta, p) + \delta e^{-(2-p)\alpha_0}}, \\
\delta_{Bk} &= \delta_{Bk}(\beta, \tau, p, \alpha_0, \delta) \\
&= \left\{\tau A_k(\beta, p) + \delta e^{-(2-p)\alpha_0}\right\}^{p-1} \left\{\tau B_k(\beta, p) + \delta e^{-(1-p)\alpha_0}\right\}^{2-p}.
\end{aligned}$$

Therefore, the prior density (17) *is conjugate.*

Proof From Lemma 1(i),

$$\begin{aligned}
L(\alpha_k - \alpha_0; p) &= u(e^{\alpha_k - \alpha_0}; 2-p) - u(e^{\alpha_k - \alpha_0}; 1-p) \\
&= e^{-(2-p)\alpha_0} u(e^{\alpha_k}; 2-p) - e^{-(1-p)\alpha_0} u(e^{\alpha_k}; 1-p) + L(-\alpha_0; p).
\end{aligned}$$

This, together with Lemma 1(ii), gives

$$\begin{aligned}
&\tau\left\{A_k(\beta, p) u(e^{\alpha_k}; 2-p) - B_k(\beta, p) u(e^{\alpha_k}; 1-p)\right\} + \delta L(\alpha_k - \alpha_0; p) \\
&= \left\{\tau A_k(\beta, p) + \delta e^{-(2-p)\alpha_0}\right\} u(e^{\alpha_k}; 2-p) \\
&\quad - \left\{\tau B_k(\beta, p) + \delta e^{-(1-p)\alpha_0}\right\} u(e^{\alpha}; 1-p) + \delta L(-\alpha_0; p) \\
&= \delta_{Bk} L(\alpha_k - \hat{\alpha}_{Bk}; p) + D_k(\beta, \tau, p, \alpha_0, \delta),
\end{aligned} \tag{19}$$

where

$$\begin{aligned}
&D_k(\beta,\,\tau,\,p,\,\alpha_0,\,\delta) \\
&= \left\{\tau A_k(\beta,\,p) + \delta e^{-(2-p)\alpha_0}\right\} u(e^{\hat\alpha_{Bk}};\,2-p) \\
&\qquad - \left\{\tau B_k(\beta,\,p) + \delta e^{-(1-p)\alpha_0}\right\} u(e^{\hat\alpha_{Bk}};\,1-p) + \delta L(-\alpha_0;\,p).
\end{aligned} \tag{20}$$

It follows from (14), (17) and (19) that

$$\begin{aligned}
&f(y_k;\,\alpha_k,\,\beta,\,\tau,\,p)\,\pi(\alpha_k-\alpha_0;\,p,\,\delta n_k) \\
&= \frac{f(y_k;\,0,\,\beta,\,\tau,\,p)}{K(p,\,\delta n_k)} \exp\left\{-\delta_{Bk} n_k L(\alpha_k-\hat\alpha_{Bk};\,p) - n_k D_k(\beta,\,\tau,\,p,\,\alpha_0,\,\delta)\right\}.
\end{aligned} \tag{21}$$

Thus, using (18), the posterior density is calculated as

$$\pi(\alpha_k-\hat\alpha_{Bk};\,p,\,\delta_{Bk}n_k) = \frac{1}{K(p,\,\delta_{Bk}n_k)} \exp\{-\delta_{Bk}n_k L(\alpha_k-\hat\alpha_{Bk};\,p)\},$$

which is of the form of the location–dispersion family (2). This completes the proof. □

Other properties of the assumed prior density (17) are given in the following Lemma, which will be applied in discussing the conjugate analysis.

Lemma 2 *Set* $\xi(p,\,t) = 1-(2-p)(p-1)\,(\partial/\partial t)\log K(p,\,t)$.

(i) Both the following are true:

$$\begin{aligned}
\mathrm{E}_\pi\left[e^{(2-p)\alpha_k}\right] &= \xi(p,\,\delta n_k)e^{(2-p)\alpha_0}, \\
\mathrm{E}_\pi\left[e^{(1-p)\alpha_k}\right] &= \xi(p,\,\delta n_k)e^{(1-p)\alpha_0},
\end{aligned}$$

where $\mathrm{E}_\pi[\,\cdot\,]$ *denotes the expectation with respect to the prior density* (17).

(ii) The Kullback-Leibler separator from $\pi(\alpha_k-\alpha_{01};\,p,\,\delta n_k)$ *to* $\pi(\alpha_k-\alpha_{02};\,p,\,\delta n_k)$ *is*

$$\mathrm{KL}(\pi(\alpha_k-\alpha_{01};\,p,\,\delta n_k),\,\pi(\alpha_k-\alpha_{02};\,p,\,\delta n_k)) = \delta n_k\xi(p,\,\delta n_k)L(\alpha_{01}-\alpha_{02}).$$

Proof

(i) The required result is obvious when $p=1$ or $p=2$ since the density (17) is a log-transformed gamma density, so suppose $1<p<2$. Differentiating both sides of the equality

$$K(p,\,t) = \int_{-\infty}^{\infty} \exp[-tL(\alpha_k-\alpha_0;\,p)]\,d\alpha_k$$

with respect to α_0 and t gives

$$\begin{aligned}
&\int_{-\infty}^{\infty} \left\{e^{(2-p)(\alpha_k-\alpha_0)} - e^{(1-p)(\alpha_k-\alpha_0)}\right\} \exp[-tL(\alpha_k-\alpha_0;\,p)]\,d\alpha_k = 0, \\
&\qquad\qquad \int_{-\infty}^{\infty} L(\alpha_k-\alpha_0;\,p)\exp[-tL(\alpha_k-\alpha_0;\,p)]\,d\alpha_k = -\frac{\partial}{\partial t}K(p,\,t).
\end{aligned}$$

Multiplying both sides of the latter equality with $(2-p)(p-1)$ and using (18),

$$\begin{aligned}
&K(p,\,t) - (2-p)(p-1)\frac{\partial}{\partial t}K(p,\,t) \\
&= \int_{-\infty}^{\infty} \left\{(p-1)e^{(2-p)(\alpha_k-\alpha_0)} - (2-p)e^{(1-p)(\alpha_k-\alpha_0)}\right\} \exp[-tL(\alpha_k-\alpha_0;\,p)]\,d\alpha_k.
\end{aligned}$$

Replace t with δn_k. Then the above two equalities form a set of linear equations

$$\mathrm{E}_\pi\left[e^{(2-p)(\alpha_k-\alpha_0)}\right] = \mathrm{E}_\pi\left[e^{(1-p)(\alpha_k-\alpha_0)}\right],$$
$$(p-1)\mathrm{E}_\pi\left[e^{(2-p)(\alpha_k-\alpha_0)}\right] + (2-p)\mathrm{E}_\pi\left[e^{(1-p)(\alpha_k-\alpha_0)}\right] = \xi(p, \delta n_k).$$

The results are the solution to this set of equations.

(ii) From Lemma 1(i),

$$L(\alpha_k - \alpha_{02}; p) - L(\alpha_k - \alpha_{01}; p)$$
$$= e^{(2-p)(\alpha_k-\alpha_{01})}u(e^{\alpha_{01}-\alpha_{02}}; 2-p) - e^{(1-p)(\alpha_k-\alpha_{01})}u(e^{\alpha_{01}-\alpha_{02}}; 1-p).$$

The definition of the Kullback-Leibler divergence, together with (i), yields the required result.

□

Now we discuss the conjugate analysis for α_k assuming on a temporary basis that β, α_0 and δ are known. As stated in Sect. 4, the loss function $L(\alpha_k - \hat{\alpha}_k; p)$ is adopted. This is a Kullback-Leibler loss function, which follows from Lemma 2(ii). The conjugate analysis of α_k is summarized in the following proposition.

Proposition 4 *A modified Pythagorean relationship*

$$\mathrm{E}_{\mathrm{post}}\Big[L(\alpha_k - \hat{\alpha}_k; p) - L(\alpha_k - \hat{\alpha}_{\mathrm{B}k}; p) - \xi(p, \delta_{\mathrm{B}k} n_k) L(\hat{\alpha}_{\mathrm{B}k} - \hat{\alpha}_k; p)\Big] = 0$$

holds for any estimator $\hat{\alpha}_k$ *where* $\mathrm{E}_{\mathrm{post}}[\cdot]$ *stands for the posterior expectation. Therefore, the estimator* $\hat{\alpha}_{\mathrm{B}k}$ *is optimal under the loss function* $L(\alpha_k - \hat{\alpha}_k; p)$.

Proof Consider the Kullback-Leibler divergence from the posterior density $\pi(\alpha_k - \hat{\alpha}_{\mathrm{B}k}; p, \delta_{\mathrm{B}k} n_k)$ to another density $\pi(\alpha_k - \hat{\alpha}_k; p, \delta_{\mathrm{B}k} n_k)$. The latter is obtained by substituting $\hat{\alpha}_{\mathrm{B}k}$ with an arbitrary estimator $\hat{\alpha}_k$. Note that the two densities have the same normalizing constant. It follows from (17) that

$$\mathrm{KL}\Big(\pi(\alpha_k - \hat{\alpha}_{\mathrm{B}k}; p, \delta_{\mathrm{B}k} n_k), \pi(\alpha_k - \hat{\alpha}_k; p, \delta_{\mathrm{B}k} n_k)\Big)$$
$$= \delta_{\mathrm{B}k} n_k \mathrm{E}_{\mathrm{post}}\Big[L(\alpha_k - \hat{\alpha}_k; p) - L(\alpha_k - \hat{\alpha}_{\mathrm{B}k}; p)\Big].$$

Apply Lemma 2(ii) to the left-hand side of this equality. □

Note that $\hat{\alpha}_{\mathrm{B}k}$ and $\delta_{\mathrm{B}k}$ (in the Bayesian context) coincide with $\hat{\alpha}_{\mathrm{M}}$ and δ_{M} (in the maximum likelihood context) respectively in Proposition 2 when δ is zero.

The family of distributions to which the conjugate prior density (17) belongs was first derived by Ohnishi & Yanagimoto [16] in the context of seeking members of the location–dispersion family having a conjugate prior. They sought location–dispersion densities $f(y-\mu)$ with a conjugate prior density of the form $\pi(\mu - m; \delta) \propto \{f(m-\mu)\}^\delta$. This requisition also yields the Normal and the von Mises distributions.

6 The optimal estimating function

We now investigate the following estimating equation of (β, τ, p):

$$\mathrm{E}_{\mathrm{post}}\left[\nabla l(\boldsymbol{y}; \boldsymbol{\alpha}, \beta, \tau, p)\right] = \mathbf{0} \tag{22}$$

where $l(\boldsymbol{y};\, \boldsymbol{\alpha},\, \beta,\, \tau,\, p) = \log f(\boldsymbol{y};\, \boldsymbol{\alpha},\, \beta,\, \tau,\, p)$ and $\nabla = (\partial/\partial\beta,\, \partial/\partial\tau,\, \partial/\partial p)^T$. This estimating equation is proved to be optimal under a certain Bayesian criterion. The hyperparameters α_0 and δ are assumed to be known in this Section.

The following proposition gives another expression of the estimating function (22). The proof is straightforward and is omitted.

Proposition 5 *It holds that*

$$\nabla \log f_{\mathrm{marg}}(\boldsymbol{y};\, \beta,\, \tau,\, p,\, \alpha_0,\, \delta) = \mathrm{E}_{\mathrm{post}}\left[\nabla l(\boldsymbol{y};\, \boldsymbol{\alpha},\, \beta,\, \tau,\, p)\right],$$

where $f_{\mathrm{marg}}(\boldsymbol{y};\, \beta,\, \tau,\, p,\, \alpha_0,\, \delta)$ *is the marginal density.*

An optimality of the estimating function (22) is shown in the following proposition. The criterion function in the proposition was adopted by Ferreira [6], which is an extended version of the one in Godambe & Kale [9] adapted to the Bayesian framework.

Proposition 6 *Suppose* α_0 *and* δ *are known, and consider an estimating function* $\boldsymbol{g}(\boldsymbol{y};\, \beta,\, \tau,\, p)$ *which is unbiased in the sense that*

$$\mathrm{E}_\pi\left[\mathrm{E}_f[\boldsymbol{g}(\boldsymbol{y};\, \beta,\, \tau,\, p)]\right] = \mathbf{0}, \tag{23}$$

where $\mathrm{E}_f[\,\cdot\,]$ *denotes the expectation with respect to the sampling density. Then, the estimating function in* (22) *is optimal with respect to the criterion*

$$\mathcal{M}[\boldsymbol{g}] = \mathrm{Tr}\,(B^{-1}A\,(B^T)^{-1}) \tag{24}$$

where $A = \mathrm{E}_\pi\left[\mathrm{E}_f[\boldsymbol{g}\boldsymbol{g}^T]\right]$ *and* $B = \mathrm{E}_\pi\left[\mathrm{E}_f[\nabla \boldsymbol{g}^T]\right]$.

Proof Since $\boldsymbol{g}(\boldsymbol{y};\, \beta,\, \tau,\, p)$ does not depend on $\boldsymbol{\alpha}$, write the unbiasedness condition (23) as

$$\mathrm{E}_{\mathrm{marg}}[\boldsymbol{g}(\boldsymbol{y};\, \beta,\, \tau,\, p)] = \mathbf{0},$$

where $\mathrm{E}_{\mathrm{marg}}[\,\cdot\,]$ is the expectation with respect to the marginal density. Similarly, the matrices A and B in the criterion function (24) can be expressed as $A = \mathrm{E}_{\mathrm{marg}}[\boldsymbol{g}\boldsymbol{g}^T]$ and $B = \mathrm{E}_{\mathrm{marg}}[\nabla \boldsymbol{g}^T]$. Using criterion (ii) in Godambe & Kale [9, Section 1.7], the optimal estimating function is $\nabla \log f_{\mathrm{marg}}(\boldsymbol{y};\, \beta,\, \tau,\, p,\, \alpha_0,\, \delta)$, which Proposition 5 proves to be equivalent to the estimating function in (22). □

An interesting relationship holds between the first element of the above optimal estimating function and the optimal estimator derived in Sect. 5. The optimal estimating function of β is expressed in terms of the optimal estimator $\hat{\alpha}_{Bk}$. The score function of β is expressed as $l_\beta(\boldsymbol{y};\, \boldsymbol{\alpha},\, \beta,\, \tau,\, p) = \sum l_{k\beta}(\boldsymbol{y}_k;\, \alpha_k,\, \beta,\, \tau,\, p)$ where

$$\begin{aligned} & l_{k\beta}(\boldsymbol{y}_k;\, \alpha_k,\, \beta,\, \tau,\, p) \\ &= -\tau \left\{ e^{(2-p)\alpha_k} \sum_{i=1}^{n_k} x_{ki} e^{(2-p)\beta x_{ki}} - e^{(1-p)\alpha_k} \sum_{i=1}^{n_k} x_{ki} y_{ki} e^{(1-p)\beta x_{ki}} \right\}. \end{aligned} \tag{25}$$

This is shown by noting that

$$\log f(\boldsymbol{y}_k;\, \alpha_k,\, \beta,\, \tau,\, p) = -\tau \sum_{i=1}^{n_k} \left\{ u\left(e^{\alpha_k + \beta x_{ki}};\, 2-p\right) - y_{ki} u\left(e^{\alpha_k + \beta x_{ki}};\, 1-p\right) \right\} + F_k,$$

where F_k is the term constant in β.

Proposition 7 *For any $k \in \{1, \ldots, K\}$,*

$$\mathrm{E}_{\mathrm{post}}\left[l_{k\beta}(\boldsymbol{y}_k;\, \alpha_k,\, \beta,\, \tau,\, p)\right] = \xi(p,\, \delta_{\mathrm{B}k} n_k)\, l_{k\beta}\left(\boldsymbol{y}_k;\, \hat{\alpha}_{\mathrm{B}k},\, \beta,\, \tau,\, p\right).$$

Therefore, the optimum estimating function is

$$\mathrm{E}_{\mathrm{post}}\left[l_{\beta}(\boldsymbol{y};\, \boldsymbol{\alpha},\, \beta,\, \tau,\, p)\right] = \sum_{k=1}^{K} \xi(p,\, \delta_{\mathrm{B}k} n_k)\, l_{k\beta}\left(\boldsymbol{y}_k;\, \hat{\alpha}_{\mathrm{B}k},\, \beta,\, \tau,\, p\right).$$

Proof Proposition 3 and Lemma 2 yield that

$$\mathrm{E}_{\mathrm{post}}\left[e^{(2-p)\alpha_k}\right] = \xi(p,\, \delta_{\mathrm{B}k} n_k) e^{(2-p)\hat{\alpha}_{\mathrm{B}k}},$$
$$\mathrm{E}_{\mathrm{post}}\left[e^{(1-p)\alpha_k}\right] = \xi(p,\, \delta_{\mathrm{B}k} n_k) e^{(1-p)\hat{\alpha}_{\mathrm{B}k}}.$$

This, together with (25), completes the proof. □

7 Estimation procedure

We propose to estimate β, τ, p and δ by maximizing the marginal likelihood function, although these parameters are assumed to be known in Sects. 4 and 5. Here set $\alpha_0 = 0$ for simplicity.

Proposition 8 *The marginal likelihood function with $\alpha_0 = 0$ is*

$$\begin{aligned} &f_{\mathrm{marg}}(\boldsymbol{y};\, \beta,\, \tau,\, p,\, 0,\, \delta) \\ &= f(\boldsymbol{y};\, \boldsymbol{0},\, \beta,\, \tau,\, p) \prod_{k=1}^{K} \left[\frac{K(p,\, \delta_{\mathrm{B}k} n_k)}{K(p,\, \delta n_k)} \exp\{-n_k D_k(\beta,\, \tau,\, p,\, 0,\, \delta)\}\right], \end{aligned}$$

where $\delta_{\mathrm{B}k} = \delta_{\mathrm{B}k}(\beta,\, \tau,\, p,\, 0,\, \delta)$ and $D_k(\beta,\, \tau,\, p,\, \alpha_0,\, \delta)$ is defined by (20).

Proof Using the expression (21) with $\alpha_0 = 0$ and (18),

$$\begin{aligned} &\int_{-\infty}^{\infty} f(\boldsymbol{y}_k;\, \alpha_k,\, \beta,\, \tau,\, p)\, \pi(\alpha_k;\, p,\, \delta n_k)\, d\alpha_k \\ &= f(\boldsymbol{y}_k;\, 0,\, \beta,\, \tau,\, p) \frac{K(p,\, \delta_{\mathrm{B}k} n_k)}{K(p,\, \delta n_k)} \exp\{-n_k D_k(\beta,\, \tau,\, p,\, 0,\, \delta)\}, \end{aligned}$$

which completes the proof. □

Our estimation procedure consists of the following two steps:

Step 1. Maximize the marginal likelihood $f_{\mathrm{marg}}(\boldsymbol{y};\, \beta,\, \tau,\, p,\, 0,\, \delta)$ with respect to β, τ, p and δ.

Step 2. Estimate α_k by plugging the estimates in Step 1 into $\hat{\alpha}_{\mathrm{B}k}(\beta,\, \tau,\, p,\, 0,\, \delta)$ in Proposition 3.

In practice, the marginal likelihood for given values of p is maximized and the maximizer $\hat{p}$ is found through a cubic spline curve computed over a suitable range.

It is interesting to compare the Bayesian estimation procedure with the maximum likelihood (ML) procedure. It follows from Proposition 2 that the ML procedure maximizes

$$f(\boldsymbol{y};\, \boldsymbol{0},\, \beta,\, \tau,\, p) \prod_{k=1}^{K} \exp\{-n_k C_k(\beta,\, \tau,\, p)\}$$

with respect to (β, τ, p). Although $\lim_{\delta \to +0} D_k(\beta, \tau, p, 0, \delta) = C_k(\beta, \tau, p)$, a positive estimate for δ results since $K(p, \delta n_k)$ tends to infinity as δ approaches to zero. Thus, our Bayesian procedure is different from the ML procedure.

An R function is available, on request, for fitting the models proposed in this paper. Evaluation of the Tweedie density function $f(y;\, \mathbf{0}, \beta, \tau, p)$ in the marginal likelihood function given in Proposition 8 is performed using numerical algorithms (Dunn & Smyth [3, 4]) as implemented in the `tweedie` package (Dunn [2]). Evaluation of the $K(p, \cdot)$ functions, defined in Eq. (18), in the marginal likelihood, is performed using the `integrate` function in R, which is based on QUADPACK routines (Piessens [17]); these programs routinely accommodate infinite limits of integration. The optimization in Step 1. is performed using constrained multivariate minimization as found in the R function `nlminb`, based on the PORT routines `http://netlib.bell-labs.com/netlib/port/`.

8 Example

Forty-five stations were selected for this study (Fig. 1), all in the same climatic region as identified by the Australian Bureau of Meteorology at `http://www.bom.gov.au/cgi-bin/climate/cgi_bin_scripts/clim_classification.cgi` (based on Gaffney [7]). All stations are considered subtropical (Cfa in the classification of Köppen [13]).

The total July rainfall from 1970 to 2006 (37 observations per station) is used here (Table 1). A plot of the rainfall at selected stations (Fig 2) shows the extreme skewness of the distributions. Also recorded, but not shown, is the average monthly southern oscillation index, or SOI [20], for the corresponding months. The SOI is defined as difference between Tahiti and Darwin air pressures, and has been linked to Australian rainfall (Stone & Auliciems [19]).

We consider a Tweedie GLM with a logarithmic link function such that $h(\mu_{ki}) = \log \mu_{ki} = \alpha_k + \beta x_{ki}$, where $x_{ki} = x_i$ represents the SOI. In this example, the intercepts α_k represent specific features of the observation stations and the slope β represents the common effect of the SOI in the region. Thus our primary interest is placed on the common slope; that is, the effect of the SOI on rainfall. The intercepts are parameters of secondary interest.

After initially using a coarse grid to determine p, the final model was fitted to the data considering a finer grid of values from $p = 1.66$ to $p = 1.73$ in steps of 0.01. A smooth cubic spline interpolation may be fitted through these computed points for a more accurate estimate. A nominal confidence interval for p is found using that $2\left[\log L(\hat{p}) - \log L(p_0)\right]$ has, asymptotically, a χ^2_1 distribution, where p_0 is the true parameter value. The profile likelihood plot (Fig. 3) indicates an estimate of $\hat{p} \approx 1.69$, with a nominal 95% confidence interval from 1.663 to 1.719 approximately. In practice, any value of p in the (say) 95% confidence interval produces very similar estimates, models and residual plots. At this empirical Bayesian estimate of p, compute $\hat{\beta} = 0.0372$, $\hat{\tau} = 0.376$ and $\hat{\delta} = 0.00173$.

For each candidate value of p, numerical methods are used to maximize the log-likelihood over the (β, δ) space, then the optimum value of τ found for each (β, δ) combination. The procedure iterates until convergence, and the entire process repeated for the next value of p under consideration.

In particular, the value of β is of interest. The profile likelihood plot for β (Fig. 4) shows a nominal 95% confidence interval for β, computed similarly to that for p, is from 0.032 to 0.0427 approximately. This interval certainly does not contain zero, so the effect of SOI is statistically significant, though the value is small so may not be of any practical significance. In this case, the maximum likelihood estimates are very similar to those computed using the empirical Bayesian approach (Table 2); this is expected since δ is small.

References

[1] CHEN, M.-H. & IBRAHIM, J. G. (2003). Conjugate priors for generalized linear models. *Statistica Sinica*, **13**, 461–476.

[2] DUNN, P. K. (2007). `tweedie`: Tweedie exponential family models, R package. Available from `http://cran.r-project.org/src/contrib/Descriptions/tweedie.html`

[3] DUNN, P. K. & SMYTH, G. K. (2005). Series evaluation of Tweedie exponential dispersion models densities. *Statistics and Computing*, **15**, 267–280.

[4] DUNN, P. K. & SMYTH, G. K. (2007). Evaluation of Tweedie exponential dispersion model densities by Fourier inversion. *Statistics and Computing*, To appear.

[5] EFRON, B. & MORRIS, C. (1973). Stein's estimation rule and its competitors—An empirical Bayes approach. *Journal of the American Statistical Association*, **68**, 117–130.

[6] FERREIRA, P. E. (1982). Estimating equations in the presence of prior knowledge. *Biometrika*, **69**, 667–669.

[7] GAFFNEY, D. O. (1971). Seasonal rainfall Zones in Australia. *Bureau of Meteorology*, Working paper No 141.

[8] GELMAN, A., CARLIN, J. B., STERN, H. S. & RUBIN, D. B. (2003). *Bayesian Data Analysis*, 2nd edn. Boca Raton: Chapman and Hall.

[9] GODAMBE, V. P. & KALE, B. K. (1991). Estimating functions: An overview. In *Estimating Functions*, ed V. P. Godambe, pp. 3–20. Oxford: Clarendon Press.

[10] IBRAHIM, J. G. & CHEN, M.-H. (2000). Power prior distributions for regression models. *Statistical Science*, **15**, 46–60.

[11] JAMES, W. & STEIN, C. (1961). Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Math. Statist. and Probability*, University of California Press, Berkeley, Vol. 1, 361–379.

[12] JØRGENSEN, B. (1997) *The Theory of Dispersion Models.* London: Chapman and Hall.

[13] KÖPPEN, W. (1931), *Grundriss der Klimakunde*, 2nd edn. Berlin: De Gruyer.

[14] MCCULLAGH, P. & NELDER, J. A. (1989). *Generalized linear models*, 2nd edn. London: Chapman and Hall.

[15] OHNISHI, T. & TSUBAKI, H. (2001). Minimization of the Fisher information matrix under a given covariance matrix function. *The Institute of Statistical Mathematics*, Research Memorandum No 819.

[16] OHNISHI, T. & YANAGIMOTO, T. (2007). Conjugate location–dispersion families. *Journal of the Japanese Statistical Society*, **37**, 307–325.

[17] PIESSENS, R. & DE DONCKER-KAPENGA, E., ÜBERHUBER, C. W. and KAHANER, D. K. (1983). *Quadpack: A Subroutine Package for Automatic Integration*, Springer-Verlag: Berlin.

[18] R DEVELOPMENT CORE TEAM (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org`

[19] STONE, R. C. & AULICIEMS, A. (1992). SOI phase relationships with rainfall in eastern Australia. *International Journal of Climatology*, **12**, 625–636.

[20] TROUP, A. J. (1965). The southern oscillation. *The Quarterly Journal of the Royal Meteorological Society*, **91**, 490–506.

[21] TSUBAKI, H. (1988). A generalization of the quasi-likelihood and its application to linear inference (in Japanese). *Doctorial dissertation, University of Tokyo.*

[22] TWEEDIE, M. C. K. (1984). An index which distinguishes between some important exponential families. Statistics: Applications and New Directions. *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference*, Indian Statistical Institute, Calcutta, pp. 579–604.

[23] YANAGIMOTO, T. & OHNISHI, T. (2005). Extensions of a conjugate prior through the Kullback-Leibler separators. *Journal of Multivariate Analysis*, **92**, 116–133.

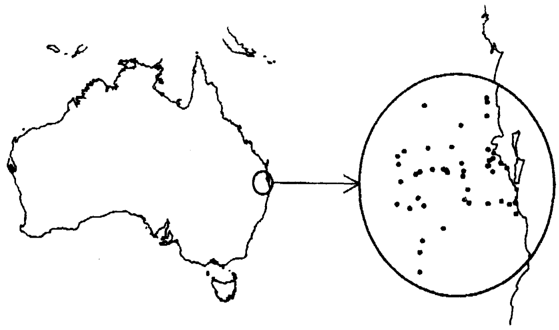

Figure 1: The location of the stations in the study

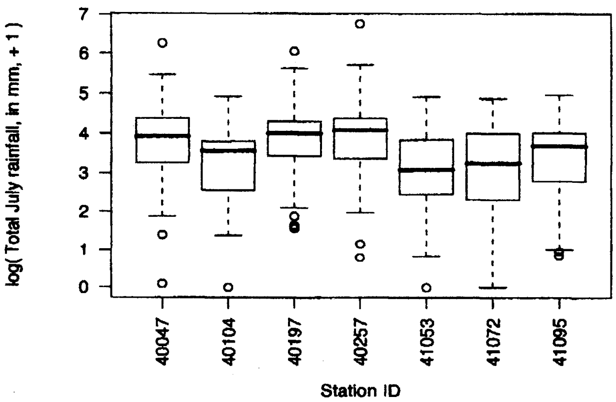

Figure 2: Boxplots of total July rainfall for selected stations

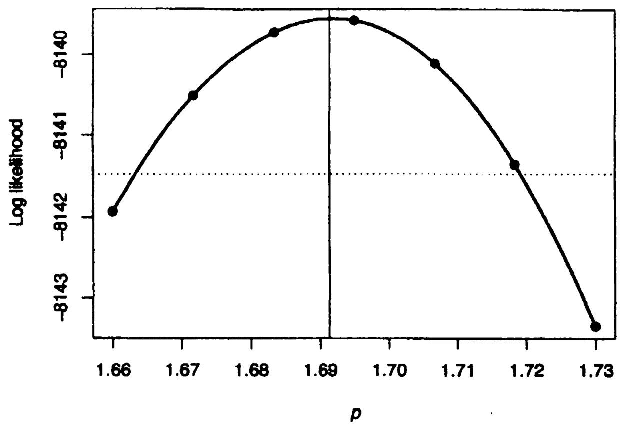

Figure 3: The profile likelihood plot for p. The points represent the actual log-likelihoods computed at the given values of p. The thick solid line is a cubic spline smooth through the computed points. The dotted horizontal line represents the height of the nominal 95% confidence interval. The vertical line is the location of the empirical Bayesian estimate of p.

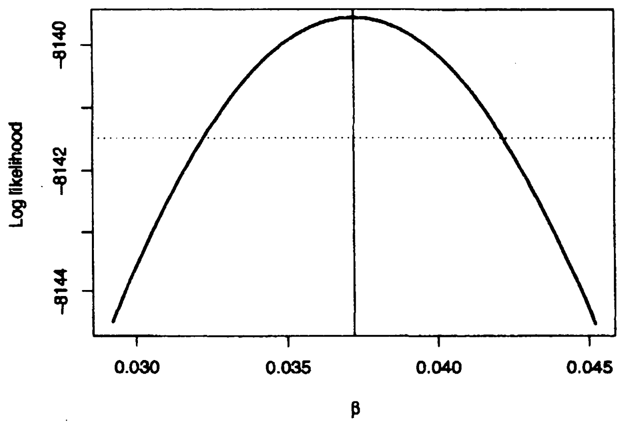

Figure 4: The profile likelihood plot for β. The thick solid line is the log-likelihood. The dotted horizontal line represents the height of the nominal 95% confidence interval. The vertical line is the location of the empirical Bayesian estimate of p.

Table 1: Summaries of the total July rainfall (in millimetres) at each station

Station	Min	5% quantile	Median	Mean	95% quantile	Max	Std dev	IQR	Percent zeros
40014	1.4	2.8	34.0	36.8	96.8	138.8	31.2	35.2	0.0
40024	0.0	0.0	29.8	35.7	99.9	159.6	35.0	39.8	8.0
40047	0.1	5.1	50.1	72.4	178.6	520.2	90.4	54.6	0.0
40075	0.0	1.6	30.7	37.7	118.3	226.6	42.8	23.9	5.0
40079	0.0	0.0	27.7	35.3	94.8	258.0	45.7	25.3	8.0
40082	0.0	0.8	29.9	36.8	89.6	306.4	52.4	28.0	3.0
40083	0.0	0.9	28.2	35.3	99.4	224.2	41.9	29.7	3.0
40094	0.8	2.1	34.5	35.2	90.2	150.8	29.7	27.8	0.0
40096	0.0	0.3	26.2	33.8	107.4	142.5	33.7	35.1	3.0
40104	0.0	0.0	34.8	35.5	107.9	137.7	33.5	32.4	8.0
40110	0.0	1.7	23.1	43.6	112.9	365.4	64.7	34.3	3.0
40117	1.0	4.0	55.4	84.9	228.4	879.9	145.5	42.7	0.0
40120	0.8	3.1	27.8	36.1	101.0	243.7	43.3	24.0	0.0
40124	0.0	0.0	24.5	34.5	114.2	204.8	40.3	29.1	10.0
40157	0.0	1.1	53.4	90.9	245.2	835.4	142.6	54.8	5.0
40158	0.0	2.1	32.2	42.8	134.1	273.3	50.4	25.7	5.0
40160	1.2	5.5	50.5	66.8	178.5	367.3	67.8	61.8	0.0
40171	0.0	2.3	32.3	53.9	137.6	481.7	80.7	32.6	3.0
40184	0.8	1.4	29.0	37.9	104.8	228.1	42.0	38.5	0.0
40190	1.0	3.5	58.8	72.7	153.9	406.2	71.1	62.0	0.0
40196	0.7	3.8	54.2	71.8	178.3	363.4	71.2	69.0	0.0
40197	3.8	4.2	54.0	69.4	184.7	427.3	79.3	44.4	0.0
40224	0.1	2.6	32.4	51.9	149.2	404.0	72.0	27.8	0.0
40229	0.2	1.4	28.0	48.6	162.2	329.6	64.3	28.6	0.0
40231	0.0	3.5	36.1	54.9	153.5	435.6	77.0	44.0	3.0
40237	2.0	6.0	30.0	54.1	161.8	430.2	76.2	28.2	0.0
40242	0.4	3.4	31.5	51.0	142.9	423.0	71.9	27.7	0.0
40244	0.2	4.9	31.2	52.5	159.4	378.8	68.8	27.3	0.0
40245	0.2	3.8	29.8	48.5	154.5	318.8	60.8	28.4	0.0
40257	1.2	5.5	58.9	88.7	234.5	867.1	145.6	51.6	0.0
40382	0.0	1.8	40.2	41.2	109.0	212.1	41.1	34.2	3.0
41001	0.0	0.1	30.4	38.8	107.0	122.8	33.6	42.4	5.0
41011	0.0	0.4	30.8	31.9	75.7	128.5	28.3	30.4	3.0
41018	0.4	0.8	31.0	34.9	87.4	115.0	28.6	37.8	0.0
41022	2.3	4.3	45.2	49.3	141.3	158.3	40.2	48.6	0.0
41053	0.0	0.0	21.1	31.5	83.3	138.4	29.8	35.8	8.0
41056	1.6	2.6	31.6	39.2	104.3	184.3	38.9	41.8	0.0
41063	0.0	0.5	30.8	37.7	93.0	131.4	31.6	39.9	3.0
41072	0.0	0.0	25.0	33.3	76.1	130.9	29.2	45.3	8.0
41082	0.0	1.2	31.2	36.1	98.6	137.4	32.6	36.2	5.0
41083	0.0	0.6	31.0	40.1	104.2	124.9	33.6	47.6	5.0
41095	1.3	1.8	39.7	45.2	128.3	143.8	36.4	39.7	0.0
41103	0.2	1.3	39.0	45.7	121.7	158.7	39.3	53.2	0.0
41116	2.0	3.1	34.4	45.0	112.9	144.1	34.4	43.4	0.0
41126	0.0	0.0	30.2	33.2	78.9	122.2	28.7	37.1	10.0

Table 2: Empirical Bayesian estimates and the maximum likelihood estimates for fitting the model to July rainfall

Parameter	Empirical Bayesian estimate	Maximum likelihood estimate
p	1.69	1.68
β	0.0372	0.0371
τ	0.376	0.369
δ	0.00173	